Data Transfer Negotiation Within the Virtual Solar Observatory

J.A. Hourclé
(GSI/NASA GSFC)

A.R. Davey
(SwRI)

I. Suarez-Sola
(NSO)

K.Q. Tian
(Stanford U.)

K. Yoshimura
(MSU-Bozeman)
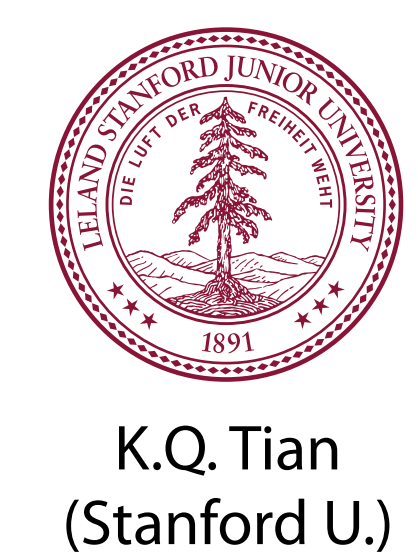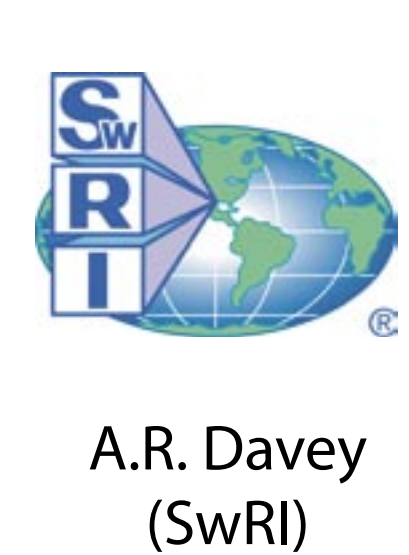
Virtual Solar Observatory

Abstract:

The Virtual Solar Observatory is attempting to build a lightweight service on top of data provider's existing architecture. This has introduced a significant level of complexity in the design of the VSO, to deal with the heterogeneous data transfer methods of the various data providers. We believe that we have an elegent solution to the problem, and can deal with the majority of transfer methods, while leaving room to expand the system as new transfer types are identified.

The Problem:

There have been a number of suggestions on how we can improve the search capabilities of the Virtual Solar Observatory in the year that it has been available to the public. The real problem, however, isn't in its searching -- it's with what you do, once you find the material that you're searching for.

Not only does each Data Provider supply their own methods of searching their archives, they also provide varied methods of obtaining data products once they have been located.

VSO attempts to solve this problem by establishing a generic interface through which users can negotiate with a Data Provider to obtain the information that may be of interest. We hope that our solution will be robust enough to handle the variety of data transfer and data requests mechanisms currently in use, while providing the flexibility necessary to support additional mechanisms in the future.

Data Transfer Negotiation:

After the user has found data products of interest, they may use VSO to gain information about obtaining the files.

In this scenario, the VSO core does not perform any manipulation of the requests, other than to act as a relay between the Data Provider. The following steps assume that the User Interface does not maintain any information about the user, and will need to prompt for any details.

1. The user interface sends VSO a list of files they are interested in, with a list of prefered data transfer methods (optional) and information about the user (optional).
2. a. If the data provider cannot support any of the data transfer methods requested, it responds with a list of supported methods.
   b. The user interface prompts the user to select a method.
   c. The user interface repeats step 1, with a data transfer method.
3. a. If the data provider requires additional information to provide that data transfer method, it responds with a list of required keywords.
   b. The user interface prompts the user to supply the necessary values for these keywords.
   c. The data provider repeats step 1, with the necessary user information.
4. The data provider responds with information detailing how to retrieve the data products.
5. The user interface (or a helper application) connects to the data archive to retrieve the data.

These steps may be reduced if the user interface uses caching of the user's preferences, or if the request can be completed without additional information. In some situations, the data request to the data archive may be triggered by the webservice, and not by the user interface.

Data Transfer Methods:

In an attempt to simplify data transfer negotiation, transfer methods were classified into four main categories, each one of which has multiple sub-categories. The system has been designed in such a way that additional categories or subcategories can be defined as they are required.

Currently, the categories are defined as follows:

URL:
The data provider will immediately supply a URL to access the information.
STAGING:
The data provider requies some time before it can provide a URL to access the information.
PUSH:
The data provider will connect to the user's system and upload the information.
OFFLINE:
The data provider will ship the information on physical media.

Sub categories define the more specific details of the transfer, such as the format in which the information may be packaged, the protocol used for data transfer, or the type of media to write the data to for offline shipment.

Users may specify multiple transfer methods that they are willing to accept, and the data provider may response with information to obtain the data using one of those methods. If there is no agreement between the user and data provider, the provider will respond with an error message containing the methods that they support.

User Information Request:

Once the user and Data Provider have agreed upon a data transfer method, the Data Provider may request additional information from the user to enable it to complete the transaction. For those transfer methods that cannot be completed immediately, a Data Provider may request an e-mail address, or other method of contacting the user for further clarification and/or notifying them when the data is available.

Some of the transfer methods may require additional information to complete their tasks – the Data Provider cannot transfer data without knowing what system to transfer the data to and what credentials to use to authenticate; they

cannot ship a CD or magtape without knowing what address to send it to.

The current implementation provides for consistent naming of information fields, to provide for caching at the user interface, or for systems to store preferences about their users, in an attempt to reduce the amount of user interaction required for downloading information.

Transfer of the Data:

The actual data transfer is a process between the Data Provider and the User, without the intervention of the VSO. This enables Data Providers to use whatever mechanisms they may already have in place to handle the transfer process.

The VSO Core does not need to know anything about the data transfer protocols that are being used, as its sole responsibility is to proxy the communication in the negotiation of the transfer. The user interface may handle the data transfer directly, or hand off the process to a helper application, such as a web browser or FTP client.

The flexibility of the system allows users to fully automate processes, such as setting up nightly tasks to find and download data products, rather than

Because the users and data providers interact directly, using existing mechanisms, the data providers may continue to use whatever existing systems they were using to capture metrics or to limit individual users without placing restrictions on all VSO users.

Future Plans:

As we begin supporting additional Data Providers with redundant Data Products, it may become worthwhile to look into handling load balancing of file downloads, in an effort to provide users with better service, while reducing the bandwidth cost to Data Providers.

As more obscure data is available through VSO, it may also becoming necessary to provide links to information about how to use or decode the data. This becoms especially important for data that were encoded using methods that require additional resources, such as calibration files, to provide meaningful information.

Future retrieval programs may be able to provide automated translation services, in order to transform the data products into an appropriate format for the user requesting the file, or the context in which is was requested.

Best Case Scenario